EXHIBIT 115

Reference Manual on Scientific Evidence

Second Edition

Federal Judicial Center 2000

There are no hard and fast rules as to which statistic is the best. In general, the bigger these measures of spread are, the more the numbers are dispersed. Particularly in small data sets, the standard deviation can be influenced heavily by a few outlying values. To remove this influence, the mean and the standard deviation can be recomputed with the outliers discarded. Beyond this, any of the statistics can be supplemented with a figure that displays much of the data.[108]

IV. What Inferences Can Be Drawn from the Data?

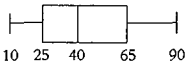
The inferences that may be drawn from a study depend on the quality of the data and the design of the study. As discussed in section II, the data might not address the issue of interest, might be systematically in error, or might be difficult to interpret due to confounding. We turn now to an additional concern—random error.[109] Are patterns in the data the result of chance? Would a pattern wash out if more data were collected?

The laws of probability are central to analyzing random error. By applying these laws, the statistician can assess the likely impact of chance error, using "standard errors," "confidence intervals," "significance probabilities," "hypothesis tests," or "posterior probability distributions." The following example illustrates the ideas. An employer plans to use a standardized examination to select trainees from a pool of 5,000 male and 5,000 female applicants. This total pool of 10,000 applicants is the statistical "population." Under Title VII of the Civil

Technically, the standard deviation is the square root of the variance; the variance is the mean square deviation from the mean. For instance, if the mean is 100, the datum 120 deviates from the mean by 20, and the square of 20 is $20^2 = 400$. If the variance (i.e., the mean of all the squared deviations) is 900, then the standard deviation is the square root of 900, that is, $\sqrt{900} = 30$. Among other things, taking the square root corrects for the fact that the variance is on a different scale than the measurements themselves. For example, if the measurements are of length in inches, the variance is in square inches; taking the square root changes back to inches.

To compare distributions on different scales, the coefficient of variation may be used: this statistic is the standard deviation, expressed as a percentage of the mean. For instance, consider the batch of numbers 1,4,4,7,9. The mean is $25/5 = 5$, the variance is $(16 + 1 + 1 + 4 + 16)/5 = 7.6$, and the standard deviation is $\sqrt{7.6} = 2.8$. The coefficient of variation is $2.8/5 = 56\%$.

108. For instance, the "five-number summary" lists the smallest value, the 25th percentile, the median, the 75th percentile, and the largest value. The five-number summary may be presented as a box plot. If the five numbers were 10, 25, 40, 65 and 90, the box plot would look like the following:



There are many variations on this idea in which the boundaries of the box, or the "whiskers" extending from it, represent slightly different points in the distribution of numbers.

109. Random error is also called sampling error, chance error, or statistical error. Econometricians use the parallel concept of random disturbance terms.

These guidelines reflect criteria proposed by the U.S. Surgeon General in 1964[112] in assessing the relationship between smoking and lung cancer and expanded upon by A. Bradford Hill in 1965.[113]

A. Is There a Temporal Relationship?

A temporal, or chronological, relationship must exist for causation. If an exposure causes disease, the exposure must occur before the disease develops.[114] If the exposure occurs after the disease develops, it cannot cause the disease. Although temporal relationship is often listed as one of many factors in assessing whether an inference of causation is justified, it is a necessary factor: Without exposure before disease, causation cannot exist.

B. How Strong Is the Association Between the Exposure and Disease?[115]

The relative risk is one of the cornerstones for causal inferences.[116] Relative risk measures the strength of the association. The higher the relative risk, the greater the likelihood that the relationship is causal.[117] For cigarette smoking, for example, the estimated relative risk for lung cancer is very high, about 10.[118] That is, the risk of lung cancer in smokers is approximately ten times the risk in nonsmokers.

A relative risk of 10, as seen with smoking and lung cancer, is so high that it is extremely difficult to imagine any bias or confounding factor that might account for it. The higher the relative risk, the stronger the association and the lower the chance that the effect is spurious. Although lower relative risks can

112. U.S. Dep't of Health, Educ., and Welfare, Public Health Serv., Smoking and Health: Report of the Advisory Committee to the Surgeon General (1964).

113. A. Bradford Hill, *The Environment and Disease: Association or Causation?*, 58 Proc. Royal Soc'y Med. 295 (1965) (Hill acknowledged that his factors could only serve to assist in the inferential process: "None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*.").

114. *See* Carroll v. Litton Sys., Inc., No. B-C-88-253, 1990 U.S. Dist. LEXIS 16833, at *29 (W.D.N.C. Oct. 29, 1990) ("[I]t is essential for . . . [the plaintiffs' medical experts opining on causation] to know that exposure preceded plaintiffs' alleged symptoms in order for the exposure to be considered as a possible cause of those symptoms . . . .").

115. Assuming that an association is determined to be causal, the strength of the association plays an important role legally in determining the specific causation question—whether the agent caused an individual plaintiff's injury. *See infra* § VII.

116. *See supra* § III.A.

117. *See* Cook v. United States, 545 F. Supp. 306, 316 n.4 (N.D. Cal. 1982); Landrigan v. Celotex Corp., 605 A.2d 1079, 1085 (N.J. 1992). The use of the strength of the association as a factor does not reflect a belief that weaker effects occur less frequently than stronger effects. *See* Green, *supra* note 39, at 652–53 n.39. Indeed, the apparent strength of a given agent is dependent on the prevalence of the other necessary elements that must occur with the agent to produce the disease, rather than on some inherent characteristic of the agent itself. *See* Rothman & Greenland, *supra* note 49, at 9–11.

118. *See* Doll & Hill, *supra* note 7.

Far more courts have confronted the role that epidemiology plays with regard to the sufficiency of the evidence and the burden of production. The civil burden of proof is described most often as requiring the fact finder to "believe that what is sought to be proved . . . is more likely true than not true."[135] The relative risk from epidemiologic studies can be adapted to this 50% plus standard to yield a probability or likelihood that an agent caused an individual's disease.[136] An important caveat is necessary, however. The discussion below speaks in terms of the magnitude of the relative risk or association found in a study. However, before an association or relative risk is used to make a statement about the probability of individual causation, the inferential judgment, described in section V, that the association is truly causal rather than spurious is required: "[A]n agent cannot be considered to cause the illness of a specific person unless

concluded that an association between tampon use and menstrually related TSS [toxic shock syndrome] cases exists."), *aff'd sub nom.* Kehm v. Procter & Gamble Mfg. Co., 724 F.2d 613 (8th Cir. 1984).

Hearsay concerns may limit the independent admissibility of the study (*see supra* note 3), but the study could be relied on by an expert in forming an opinion and may be admissible pursuant to Fed. R. Evid. 703 as part of the underlying facts or data relied on by the expert.

In *Ellis v. International Playtex, Inc.*, 745 F.2d 292, 303 (4th Cir. 1984), the court concluded that certain epidemiologic studies were admissible despite criticism of the methodology used in the studies. The court held that the claims of bias went to the studies' weight rather than their admissibility. *Cf.* Christophersen v. Allied-Signal Corp., 939 F.2d 1106, 1109 (5th Cir. 1991) ("As a general rule, questions relating to the bases and sources of an expert's opinion affect the weight to be assigned that opinion rather than its admissibility . . . . "), *cert. denied*, 503 U.S. 912 (1992).

134. Even if evidence is relevant, it may be excluded if its probative value is substantially outweighed by prejudice, confusion, or inefficiency. Fed. R. Evid. 403. However, exclusion of an otherwise relevant epidemiologic study on Rule 403 grounds is unlikely.

In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 591 (1993), the Court invoked the concept of "fit," which addresses the relationship of an expert's scientific opinion to the facts of the case and the issues in dispute. In a toxic substance case in which cause in fact is disputed, an epidemiologic study of the same agent to which the plaintiff was exposed that examined the association with the same disease from which the plaintiff suffers would undoubtedly have sufficient "fit" to be a part of the basis of an expert's opinion. The Court's concept of "fit," borrowed from *United States v. Downing*, 753 F.2d 1224, 1242 (3d Cir. 1985), appears equivalent to the more familiar evidentiary concept of probative value, albeit one requiring assessment of the scientific reasoning the expert used in drawing inferences from methodology or data to opinion.

135. 2 Edward J. Devitt & Charles B. Blackmar, Federal Jury Practice and Instruction § 71.13 (3d ed. 1977); *see also* United States v. Fatico, 458 F. Supp. 388, 403 (E.D.N.Y. 1978) ("Quantified, the preponderance standard would be 50%+ probable."), *aff'd*, 603 F.2d 1053 (2d Cir. 1979), *cert. denied*, 444 U.S. 1073 (1980).

136. An adherent of the frequentist school of statistics would resist this adaptation, which may explain why so many epidemiologists and toxicologists also resist it. To take the step identified in the text of using an epidemiologic study outcome to determine the probability of specific causation requires a shift from a frequentist approach, which involves sampling or frequency data from an empirical test, to a subjective probability about a discrete event. Thus, a frequentist might assert, after conducting a sampling test, that 60% of the balls in an opaque container are blue. The same frequentist would resist the statement, "The probability that a single ball removed from the box and hidden behind a screen is blue is 60%." The ball is either blue or not, and no frequentist data would permit the latter statement. "[T]here is no logically rigorous definition of what a statement of probability means with reference to an individual instance . . . ." Lee Loevinger, *On Logic and Sociology*, 32 Jurimetrics J. 527, 530 (1992); *see*

it is recognized as a cause of that disease in general."[137] The following discussion should be read with this caveat in mind.[138]

The threshold for concluding that an agent was more likely than not the cause of an individual's disease is a relative risk greater than 2.0. Recall that a relative risk of 1.0 means that the agent has no effect on the incidence of disease. When the relative risk reaches 2.0, the agent is responsible for an equal number of cases of disease as all other background causes. Thus, a relative risk of 2.0 (with certain qualifications noted below) implies a 50% likelihood that an exposed individual's disease was caused by the agent. A relative risk greater than 2.0 would permit an inference that an individual plaintiff's disease was more likely than not caused by the implicated agent.[139] A substantial number of courts in a variety of toxic substances cases have accepted this reasoning.[140]

also Steve Gold, Note, *Causation in Toxic Torts: Burdens of Proof, Standards of Persuasion and Statistical Evidence*, 96 Yale L.J. 376, 382–92 (1986). Subjective probabilities about discrete events are the product of adherents to Bayes Theorem. *See* Kaye, *supra* note 67, at 54–62; David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.D, in this manual.

137. Cole, *supra* note 53, at 10284.

138. We emphasize this caveat, both because it is not intuitive and because some courts have failed to appreciate the difference between an association and a causal relationship. *See, e.g.*, Forsyth v. Eli Lilly & Co., Civ. No. 95-00185 ACK, 1998 U.S. Dist. LEXIS 541, at *26–*31 (D. Haw. Jan. 5, 1998). *But see* Berry v. CSX Transp., Inc., 709 So. 2d 552, 568 (Fla. Dist. Ct. App. 1998) ("From epidemiological studies demonstrating an association, an epidemiologist may or may not infer that a causal relationship exists.").

139. *See* Davies v. Datapoint Corp., No. 94-56-P-DMC, 1995 U.S. Dist. LEXIS 21739, at *32–*35 (D. Me. Oct. 31, 1995) (holding that epidemiologist could testify about specific causation, basing such testimony on the probabilities derived from epidemiologic evidence).

140. *See* DeLuca v. Merrell Dow Pharms., Inc., 911 F.2d 941, 958–59 (3d Cir. 1990) (Bendectin allegedly caused limb reduction birth defects); *In re* Joint E. & S. Dist. Asbestos Litig., 964 F.2d 92 (2d Cir. 1992) (relative risk less than 2.0 may still be sufficient to prove causation); Daubert v. Merrell Dow Pharms., Inc., 43 F.3d 1311, 1320 (9th Cir.) (requiring that plaintiff demonstrate a relative risk of 2), *cert. denied*, 516 U.S. 869 (1995); Pick v. American Med. Sys., Inc., 958 F. Supp. 1151, 1160 (E.D. La. 1997) (recognizing that a relative risk of 2 implies a 50% probability of specific causation, but recognizing that a study with a lower relative risk is admissible, although ultimately it may be insufficient to support a verdict on causation); Sanderson v. International Flavors & Fragrances, Inc., 950 F. Supp. 981, 1000 (C.D. Cal. 1996) (acknowledging a relative risk of 2 as a threshold for plaintiff to prove specific causation); Manko v. United States, 636 F. Supp. 1419, 1434 (W.D. Mo. 1986) (swine flu vaccine allegedly caused Guillain-Barré syndrome), *aff'd in part*, 830 F.2d 831 (8th Cir. 1987); Marder v. G.D. Searle & Co., 630 F. Supp. 1087, 1092 (D. Md. 1986) (pelvic inflammatory disease allegedly caused by Copper 7 IUD), *aff'd without op. sub nom.* Wheelahan v. G.D. Searle & Co., 814 F.2d 655 (4th Cir. 1987); *In re* "Agent Orange" Prod. Liab. Litig., 597 F. Supp. 740, 835–37 (E.D.N.Y. 1984) (Agent Orange allegedly caused a wide variety of diseases in Vietnam veterans and their offspring), *aff'd*, 818 F.2d 145 (2d Cir. 1987), *cert. denied*, 484 U.S. 1004 (1988); Cook v. United States, 545 F. Supp. 306, 308 (N.D. Cal. 1982) (swine flu vaccine allegedly caused Guillain-Barré syndrome); Landrigan v. Celotex Corp., 605 A.2d 1079, 1087 (N.J. 1992) (relative risk greater than 2.0 "support[s] an inference that the exposure was the probable cause of the disease in a specific member of the exposed population"); Merrell Dow Pharms., Inc. v. Havner, 953 S.W.2d 706, 718 (Tex. 1997) ("The use of scientifically reliable epidemiological studies and the requirement of more than a doubling of the risk strikes a balance between the needs of our legal system and the limits of science."). *But cf. In re* Fibreboard Corp., 893 F.2d 706, 711–12 (5th Cir. 1990) (The court disapproved a trial in which several representative

exposure to the agent caused the plaintiff's disease. Similarly, an individual plaintiff may be able to rule out other known (background) causes of the disease, such as genetics, that increase the likelihood that the agent was responsible for that plaintiff's disease. Pathological-mechanism evidence may be available for the plaintiff that is relevant to the cause of the plaintiff's disease.[145] Before any causal relative risk from an epidemiologic study can be used to estimate the probability that the agent in question caused an individual plaintiff's disease, consideration of these (and similar) factors is required.[146]

Having additional evidence that bears on individual causation has led a few courts to conclude that a plaintiff may satisfy his or her burden of production even if a relative risk less than 2.0 emerges from the epidemiologic evidence.[147] For example, genetics might be known to be responsible for 50% of the incidence of a disease independent of exposure to the agent.[148] If genetics can be ruled out in an individual's case, then a relative risk greater than 1.5 might be sufficient to support an inference that the agent was more likely than not responsible for the plaintiff's disease.[149]

145. *See* Tobin v. Astra Pharm. Prods., Inc., 993 F.2d 528 (6th Cir.) (plaintiff's expert relied predominantly on pathogenic evidence), *cert. denied*, 510 U.S. 914 (1993).

146. *See* Merrell Dow Pharms., Inc. v. Havner, 953 S.W.2d 706, 720 (Tex. 1997); Mary Carter Andrues, Note, *Proof of Cancer Causation in Toxic Waste Litigation*, 61 S. Cal. L. Rev. 2075, 2100–04 (1988). An example of a judge sitting as fact finder and considering individual factors for a number of plaintiffs in deciding cause in fact is contained in *Allen v. United States*, 588 F. Supp. 247, 429–43 (D. Utah 1984), *rev'd on other grounds*, 816 F.2d 1417 (10th Cir. 1987), *cert. denied*, 484 U.S. 1004 (1988); *see also* Manko v. United States, 636 F. Supp. 1419, 1437 (W.D. Mo. 1986), *aff'd*, 830 F.2d 831 (8th Cir. 1987).

147. *See, e.g.*, Grassis v. Johns-Manville Corp., 591 A.2d 671, 675 (N.J. Super. Ct. App. Div. 1991): "The physician or other qualified expert may view the epidemiological studies and factor out other known risk factors such as family history, diet, alcohol consumption, smoking . . . or other factors which might enhance the remaining risks, even though the risk in the study fell short of the 2.0 correlation." *See also In re* Joint E. & S. Dist. Asbestos Litig., 52 F.3d 1124 (2d Cir. 1995) (holding that plaintiff could provide sufficient evidence of causation without proving a relative risk greater than 2); *In re* Joint E. & S. Dist. Asbestos Litig., 964 F.2d 92, 97 (2d Cir. 1992), *rev'g* 758 F. Supp. 199, 202–03 (S.D.N.Y. 1991) (requiring relative risk in excess of 2.0 for plaintiff to meet burden of production); Jones v. Owens-Corning Fiberglas Corp., 672 A.2d 230 (N.J. Super. Ct. App. Div. 1996).

148. *See In re* Paoli R.R. Yard PCB Litig., 35 F.3d 717, 758–59 (3d Cir. 1994) (discussing the technique of differential diagnosis to rule out other known causes of a disease for a specific individual).

149. The use of probabilities in excess of .50 to support a verdict results in an all-or-nothing approach to damages that some commentators have criticized. The criticism reflects the fact that defendants responsible for toxic agents with a relative risk just above 2.0 may be required to pay damages not only for the disease that their agents caused, but also for all instances of the disease. Similarly, those defendants whose agents increase the risk of disease by less than a doubling may not be required to pay damages for any of the disease that their agents caused. *See, e.g.*, 2 American Law Inst., Reporter's Study on Enterprise Responsibility for Personal Injury: Approaches to Legal and Institutional Change 369–75 (1991). To date, courts have not adopted a rule that would apportion damages based on the probability of cause in fact in toxic substances cases.

including such details as the number of animals per cage, dose and chemical verification, and the handling of tissue specimens. GLP practices are remarkably similar across agencies, but the tests called for differ depending on mission. For example, there are major differences between the FDA's and the EPA's required procedures for testing drugs and environmental chemicals.[30] The FDA requires and specifies both efficacy and safety testing of drugs in humans and animals. Carefully controlled clinical trials using doses within the expected therapeutic range are required for premarket testing of drugs because exposures to prescription drugs are carefully controlled and should not exceed specified ranges or uses. However, for environmental chemicals and agents, no premarket testing in humans is required by the EPA. Moreover, since exposures are less predictable, a wider range of doses usually is given in the animal tests.[31]

Since exposures to environmental chemicals may continue over the lifetime and affect both young and old, test designs called lifetime bioassays have been developed in which relatively high doses are given to experimental animals. Interpretation of results requires extrapolation from animals to humans, from high to low doses, and from short exposures to multiyear estimates. It must be emphasized that less than 1% of the 60,000–75,000 chemicals in commerce have been subjected to a full safety assessment, and there are significant toxicological data on only 10%–20%.

Risk assessment is an approach increasingly used by regulatory agencies to estimate and compare the risks of hazardous chemicals and to assign priority for avoiding their adverse effects.[32] The National Academy of Sciences defines four components of risk assessment: hazard identification, dose–response estimation, exposure assessment, and risk characterization.[33]

Although risk assessment is not an exact science, it should be viewed as a

safety of consumer products is described in *United States v. Keplinger*, 776 F.2d 678 (7th Cir. 1985), *cert. denied*, 476 U.S. 1183 (1986). Keplinger and the other defendants in this case were toxicologists who were convicted of falsifying data on product safety by underreporting animal morbidity and mortality and omitting negative data and conclusions from their reports.

30. *See, e.g.*, 40 C.F.R. §§ 160, 792 (1993); Lu, *supra* note 14, at 89.

31. It must be appreciated that the development of a new drug inherently requires searching for an agent that at useful doses has a biological effect (e.g., decreasing blood pressure), whereas those developing a new chemical for consumer use (e.g., a house paint) hope that at usual doses no biological effects will occur. There are other compounds, such as pesticides and antibacterial agents, for which a biological effect is desired, but it is intended that at usual doses humans will not be affected. These different expectations are part of the rationale for the differences in testing information available for assessing toxicological effects.

32. Committee on Risk Assessment Methodology, National Research Council, *supra* note 19, at 1.

33. *See generally* National Research Council, Risk Assessment in the Federal Government: Managing the Process (1983); Bernard D. Goldstein, *Risk Assessment/Risk Management Is a Three-Step Process: In Defense of EPA's Risk Assessment Guidelines*, 7 J. Am. C. Toxicol. 543 (1988); Bernard D. Goldstein, *Risk Assessment and the Interface Between Science and Law*, 14 Colum. J. Envtl. L. 343 (1989).